CS 134: Networks Section 7

Prof. Yaron Singer Spring 2017

1 Adaptive Seeding - Motivation

Adaptive seeding is a model introduced to take advantage of the Friendship Paradox in social network models in order to construct better algorithms for influence maximization. This influence maximization model is found in many places - for example, in marketing, online retailers are only able to reach those that are subscribed to their mailing list or have visited their online store.

Simply maximizing influence among the core sample is a terrible approach - given that social networks are often distributed heavy-tail, influence among the core sample is likely to be low, and thus even an optimal maximization among the core sample is likely to be far from the global optimum.

Adaptive seeding is a solution that takes instead a two-stage approach. Given some budget, one first spends some of it on some users so they invite their own friends to join, then spends the rest on the maximal influencers in the group of people that then join. This leverages the Friendship Paradox - among the core sample, they're in general likely to be of low influence, but they are very likely to have friends that have high influence. The first stage then encourages these high influence friends to join, at which point we spend the rest of the budget on them in order to maximize our influence.

We now turn to a recap of influence maximization and the Friendship Paradox before moving into a technical investigation of adaptive seeding.

2 Recap: Influence Maximization

Recall the problem of influence maximization, which aims to find the best nodes in a network to "seed" in order to propagate some content, idea, or "infection". In general, the problem states for an input:

- A graph $G = (V, E)$,
- a limit $k \in \mathbb{N}$, and
- an influence function $f : 2^V \Rightarrow \mathbb{R}$ mapping subsets $S \subseteq V$ to their influence value $f(S)$,

what is the maximum influence we can obtain across subsets of size at most k, i.e.:

$$\max_{|S| \leq k} f(S)$$

Note that in general, we cannot simply check every possible subset of V of size at most k and attempt to maximize across these; this is both unlikely and prohibitively computationally expensive. Instead, we aim to come within an approximation of the optimal $f^*(S)$, using a variety of strategies. We have shown already that additive and coverage functions are monotone submodular, and for such monotone submodular functions, the "greedy" approach is guaranteed to come with $(1 - \frac{1}{e})$ of the optimal value. However, this is only 63%, and is still prohibitively expensive! Can we do better?

The greedy algorithm and several definitions are quoted again below for your convenience.

Algorithm 1 Greedy Algorithm

```
Set $S = \emptyset$
while $|S| \leq k$ do
    $S \leftarrow S \cup \text{argmax}_{a \in N} f_S(a)$
end while
return $S$
```

Definition. *A function $f : 2^N \to \mathbb{R}$, is* ***monotone*** *if $S \subseteq T \implies f(S) \leq f(T)$.*

Definition. *Given a function $f : 2^N \to \mathbb{R}$, the* ***marginal contribution*** *of an element $e \in N$ to $S \subseteq N$ is $f_S(e) = f(S \cup e) - f(S)$.*

Definition. *A function $f : 2^N \to \mathbb{R}$ is* ***submodular*** *if for any $e \in N$ we have that:*

$$f_S(e) \geq f_T(e) \;\; \forall S \subseteq T \subseteq N$$

Definition. *A function $f : 2^N \to \mathbb{R}$ is* ***subadditive*** *if for any $S, T \subseteq N$ we have that:*

$$f(S \cup T) \leq f(S) + f(T)$$

3 Recap: Friendship Paradox

To find inspiration, we go back to the Friendship Paradox. Recall that informally, the Friendship Paradox states that on average, your friends are more "influential" than you are, where in this case influence is measured by degree. Formally, given a graph $G = (V, E)$, G has the Friendship Paradox property if:

$$P(d(n) \leq \frac{\sum_{i \in \mathbb{N}(n)} i}{d(i)}$$

is relatively high. Alternatively, this may be formulated as the expected ratio between $\sum_{i \in \mathbb{N}(n)} i$ and $d(i)$ across all nodes i in the graph G. Note that this property is hard to find in graphs in general; however, under certain models of social networks, this property arises quite often, and its existence has been empirically verified in several large online networks.

3.1 Feld's Soft Version

Feld (1991) showed a soft version of the Friendship Paradox to be unilaterally true across any graph, in that the average degree of a node is less than or greater than the expected degree of a randomly selected neighbor over all neighbors in the graph. This is quick to show. Let μ be the average degree of nodes in the graph, i.e.

$$\mu = \frac{\sum_{i=1}^{n} d(i)}{n}$$

The average degree of neighbors in the graph is:

$$\mu' = \frac{\sum_{i=1}^{n} \text{total degree of } \mathbb{N}(i)}{\sum_{i=1}^{n} d(i)}$$

For any node i, its degree $d(i)$ is considered $d(i)$ times, since it's connected to that many nodes and is thus those nodes' neighbor. Therefore, each node i contributes $d(i)^2$ to the numerators and $d(i)$ to the denominator, i.e. the above is equal to:

$$\mu' = \frac{\sum_{i=1}^{n} d(i)^2}{\sum_{i=1}^{n} d(i)}$$

It can be shown after some quick algebra that this is just:

$$\mu' = \frac{\sigma^2}{\mu} + \mu$$

Thus, μ' is greater than or equal to μ in any graph. However, this "soft" version of the Friendship Paradox is very, very different from the Paradox. To see this, consider a star graph with n nodes. The "soft" μ' is $\frac{n}{2}$, whereas the actual μ' under the Friendship Paradox interpretation is $n - 2 + \frac{2}{n}$. Perhaps more strikingly, construct from any graph G the "clique-ified" graph G' seen in previous psets, where each node i is expanded into a regular complete sub-graph of size $i + 1$. Note that the Friendship Paradox is nonexistent here! However, Feld's calculations would still be nonzero. Evidently there is a massive difference.

3.2 Social Network Models and the Friendship Paradox

In general, we are interested in social network models that have some constant probability that a random node satisfies the Friendship Paradox. A common assumption about these social networks is that they follow a heavy-tailed distribution, such as the power law distribution we have been using so much in this course. (Some others are log-normal and double Pareto distributions). However, power-law distributions do not in general have the Friendship Paradox property; in fact, "clique-ified" graphs have power-law distributions, but the Friendship Paradox property does not apply. We thus construct the Perturbed Power Law Graph model:

Definition 1. *Given a graph G whose degree distribution is power law and some probability $p \in [0,1]$, its **perturbed power law graph** $G(p)$ is the graph G where every one of the edges in G is rewired with probability p.*

Note that "rewiring" entails removing the edges and randomly reconnecting all of the remaining stubs at once. Lattanzi and Singer (2015) showed that the Friendship Paradox applies here, i.e.:

Theorem 2. *([Lattanzi and Singer 2015]). For any perturbed power-law graph with constant probability there is an asymptotic gap between the average degree of a random set of poloylogarithmic size and the average degree of its set of neighbors.*

Note that the above does NOT prove that the paradox exists for a single node; rather, it shows that for samples of polylogarithmic size, the paradox applies as a whole. However, it does still imply that seeding a polylogarithmically sized sample set's neighbors yields a significant advantage. Thus, given our core set, if with constant probability a sample within the core set is connected to a set of size on the same order that can reach asymptotically more nodes, then the adaptive seeding algorithm may be able to reach an asymptotically larger number of nodes than by applying influence maximization on the sample alone.

4 Adaptive Seeding - Technical

We are given some core sample X of a network, their neighbors $N(X)$, a set of probabilities p_i associated with each neighbor, a budget $k \in \mathbb{N}$, and an influence function $f : 2^{N(X)} \Rightarrow \mathbb{R}$. In the first stage, we choose a subset $S \subseteq X$, using $|S|$ of the budget, and causing each one of the neighbors in $N(S)$ to realize with probability p_i. In the second stage, we spend the remaining $k - |S|$ of the budget to optimize the influence function f over the realized neighbors. The goal is then to select a subset $S \subseteq X$ that allows us to maximize f in expectation over all realizations of the neighbors $N(S)$ with the remaining budget. Mathematically, the objective is:

$$\max_{S \subseteq X} \sum_{i=1}^{m} f(T_i) \cdot p(R_i)$$

$$T_i \subseteq R_i \cap N(S) \ \ \forall i \in [m]$$

$$|S| + |T_i| \leq k \ \ \forall i \in [m]$$

It can be shown that this is within $(1 - \frac{1}{e})$ of the optimal solution across the entire network (with some small assumptions).

Exercise Prove this.

One common use of the adaptive seeding model is when we are considering an additive influence function, i.e. $f(S) = \sum_{a \in S} w_a$, where w_a is the weight of the node a under this function f (or, say, the influence of a). This then reduces to maximizing a monotone submodular function under the cardinality limit k using a two-stage algorithm.

We first show that this function is monotone submodular. Define

$$F_t(S) = \sum_{i=1}^{m} p(R_i) \left(\max_{|T| \leq t, T \subseteq S \cap R_i} f(T) \right)$$

Or the expected maximum influence exerted by any subset of size at most t among set S. Note that this is just a positive weighted sum of monotone submodular functions[1], and thus $F_t(S)$ is also monotone submodular.

The optimal solution under the adaptive seeding model has some partition $(k - t, t)$ such that it spends $k - t$ on the core set and t on the neighbors, and thus, running a greedy algorithm across all functions F_t for t from 1 to $k - 1$ and taking the maximum value across all of these gets us a value within $(1 - \frac{1}{e})$ of the optimal solution.

[1] To see that $F_t(S) = \max_{|T| \leq t, T \subseteq N(S)} f(T)$ is monotone submodular when f is additive, it suffices to show that the marginal contribution of another element a is at most the marginal contribution of a to any superset of T. This can be easily seen by considering $f(S \cup a), f(S)$, and looking at what happens when we inject new elements from T/S into these sets.